NORMALIZED CUTS ARE APPROXIMATELY INVERSE EXIT TIMES

MATAN GAVISH\(^\ast\) AND BOAZ NADLER\(^\dagger\)

Abstract. The Normalized Cut is a widely used measure of separation between clusters in a graph. In this paper we provide a novel probabilistic perspective on this measure. We show that for a partition of a graph into two weakly connected sets, \( V = A \cup B \), in fact \( \text{Ncut}(V) = 1/\tau_{A \to B} + 1/\tau_{B \to A} \), where \( \tau_{A \to B} \) is the uni-directional characteristic exit time of a random walk from subset \( A \) to subset \( B \). Using matrix perturbation theory, we show that this approximation is exact to first order in the connection strength between the two subsets \( A \) and \( B \), and derive an explicit bound for the approximation error. Our result implies that for a Markov chain composed of a reversible component \( A \) that is weakly connected to an absorbing cluster \( B \), the easy-to-compute Normalized Cut measure is a reliable proxy for the chain’s spectral gap.

Key words. normalized cut, graph partitioning, Markov chain, characteristic exit time, generalized eigenvalues, perturbation bounds, matrix perturbation theory.

AMS subject classifications. 62H30, 15A18, 15A42, 60J20, 60J10, 65H17.

1. Introduction. Clustering of data is a fundamental problem in many fields. When the data is represented as a symmetric weighted similarity graph, clustering becomes a graph partitioning problem. In this paper we focus on the popular normalized cut criteria for graph partitioning. The normalized cut (Ncut) criteria, and its variants based on conductance, are widely used measures of separation between clusters in a graph in a variety of applications [7, 17]. Since optimizing the Ncut measure is in general an NP-hard problem, several researchers suggested various spectral-based approximation methods, see for example [3, 15]. These spectral methods, in turn, have interesting connections to kernel K-means clustering and to non-negative matrix factorizations, see [4, 5].

In this paper we focus on the normalized cut measure itself. Similar to previous works [10, 14], we consider a random walk on the weighted graph, with the probability to jump from one node to another proportional to the edge weight between them. Our contribution is a novel Markov chain based perspective on this measure. Our main result is that for a partition of a graph \( V \) into two disjoint sets, \( V = A \cup B \),

\[
\text{Ncut}(V) \approx \frac{1}{\tau_{A \to B}} + \frac{1}{\tau_{B \to A}}
\]

(1.1)

where \( \tau_{A \to B} \) is the characteristic exit time of a random walk from cluster \( A \) to cluster \( B \), as defined explicitly below. Furthermore, we provide an explicit bound for the approximation error in Eq.(1.1), which is second order in the connection strengths between the two subsets \( A \) and \( B \). Our proof is based on matrix perturbation theory for generalized eigenvalue problems [18]. Interestingly, (1.1) can be viewed as a discrete analog of a well-known result for continuous diffusion processes with two metastable states [9], see Section 4.

From a mathematical point of view, when the connection strengths between the two clusters are small, the two clusters become nearly uncoupled and the resulting Markov chain nearly decomposable. The properties of such singularly perturbed Markov chains have been a subject of several works, see the surveys [16, 1] and the many references therein. Some works considered the behavior of the stationary distribution, others considered mean first passage times, typically focusing on singular Laurent series expansions, and yet others considered the important problem of dimensionality reduction and aggregation of states for such nearly reducible Markov chains, see for example [6, 11]. Our focus, in contrast, is on the relation of such Markov chains to the

\( ^\ast\)Department of Statistics, Stanford University, Stanford, 94305, CA (gavish@stanford.edu)

\( ^\dagger\)Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, 76100, Israel (boaz.nadler@weizmann.ac.il)
normalized cut measure - an easy-to-compute quantity used in a variety of applied scenarios. To this end, our analysis involves the study of a singularly perturbed Markov chain with transient states, as we consider one of the clusters as absorbing with respect to a random walk that starts at the other cluster. Furthermore, rather than series expansions, we derive explicit error bounds for the approximation in (1.1).

Our result has several implications and potential applications. First, in the context of clustering, Eq. (1.1) provides a precise probabilistic statement regarding the relation between the algebraic, easy-to-compute Ncut measure of a given partition, and the eigenvalues corresponding to local random walks on each of these partitions. Our result is different from previously derived global spectral bounds that depend on the eigenvalues of the random walk on the entire graph. In particular, our analysis implies that for weakly connected clusters, the normalized cut measure essentially considers only the time it takes for a random walk to exit a given cluster, without taking into account its internal structure. As discussed in more detail in [12], this leads to a fundamental limitation of the Ncut measure to detect structures at different scales in a large graph, and motivates the use of other measures that also take into account the internal characteristics of potential clusters [13].

Another potential application of our result is in the study of stochastic dynamical systems and in particular their discretization into reversible finite Markov chains [3]. In this field, a fundamental problem is the identification of almost invariant sets, e.g. subsets of phase space where a stochastic system stays for a long time before exiting. Our observation above, in the context of clustering, is also relevant here. As complex dynamical systems may exhibit invariant sets at very different spatial and time scales, our analysis suggests that the Ncut measure or its various spectral relaxations may not be appropriate in such cases.

Yet another application of our result is to the numerical evaluation of the spectral gap of nearly decomposable Markov matrices. In the study of stochastic dynamical systems, the equivalent quantity of interest are the mean exit times from given or a-priori known metastable sets, and their dependence on various parameters, such as temperature, bond-strengths etc. Upon discretization of such systems, this amounts to the exit times of nearly uncoupled Markov chains. In this context, one may apply Eq. (1.1) in the reverse direction: Rather than computing the second largest eigenvalue of a potentially very large matrix, and from it the characteristic exit time from this metastable state, one may approximate it by the easy-to-compute normalized cut measure, with explicit guarantees as to its accuracy.

The paper is organized as follows. The normalized cut and relevant notation is introduced in Section 2. Our main result is stated in Section 3, discussed in Section 4, and demonstrated using a simple numerical simulation in Section 5. The proof appears in Section 6, with some auxiliary results deferred to the Appendix.

2. Preliminaries. Let \((V, W)\) be a connected weighted graph with node set \(V = \{1, \ldots, n\}\) and an \(n \times n\) affinity (or edge weight) matrix \(W = (W_{i,j})_{1 \leq i, j \leq n}\). The matrix \(W\) is assumed to be symmetric with non-negative entries, and such that \(W_{ii} > 0\) for all \(i = 1, \ldots, n\).

2.1. The normalized cut of a graph partition. Consider a partition of the node set into two disjoint "clusters", \(V = A \cup B\). Without loss of generality, we denote the clusters by \(A = \{1, \ldots, k\}\) and \(B = \{k + 1, \ldots, n\}\). As discussed in the introduction, the normalized cut [17] is a widely used, easy to calculate measure of separation for a given graph partition:

**Definition 2.1.** For a partition \(V = A \cup B\), define

\[
\text{Cut} (A, B) = \sum_{i \in A, j \in B} W_{i,j}
\]

\[
\text{Assoc} (A, V) = \sum_{i \in A, j \in V} W_{i,j}
\]
Then the normalized cut of $A$ with respect to $B$ is

$$N\text{cut}(A, B) = \frac{\text{Cut}(A, B)}{\text{Assoc}(A, V)} \tag{2.1}$$

whereas the multiway normalized cut of the partition $V = A \uplus B$ is defined as $[8, 19]$

$$MN\text{cut}(A \uplus B) = N\text{cut}(A, B) + N\text{cut}(B, A) = \frac{\text{Cut}(A, B)}{\text{Assoc}(A, V)} + \frac{\text{Cut}(B, A)}{\text{Assoc}(B, V)}. \tag{2.2}$$

The goal of this paper is to provide a precise probabilistic interpretation of this measure, which holds when the clusters $A$ and $B$ are “well separated” in a sense defined below.

2.2. Notation. For any $i \in V$, denote by

$$W^i,A = \sum_{j \in A} W_{i,j}, \quad W^i,B = \sum_{j \in B} W_{i,j} \tag{2.3}$$

the total affinity between the node $i$ and the clusters $A$ and $B$, respectively. Similarly, denote

$$W^{A,A} = \sum_{i \in A} W^i,A, \quad W^{A,B} = \sum_{i \in A} \sum_{j \in A} W_{i,j}, \quad W^{A,B} = \sum_{i \in A} \sum_{j \in B} W_{i,j}. \tag{2.4}$$

In this notation, the normalized cut measure (2.1) is

$$N\text{cut}(A, B) = \frac{W^{A,B}}{W^{A,A} + W^{A,B}}. \tag{2.5}$$

With an eye towards a matrix perturbation approach, we define the connectivity parameter

$$\varepsilon = \max_{i \in A} \frac{W_{i,B}}{W_{i,A}}. \tag{2.6}$$

In this paper we focus on situations where $W_{i,B} \ll W_{i,A}$ for all $i \in A$, namely all nodes $i \in A$ are weakly connected to $B$. Equivalently, $\varepsilon \ll 1$, which allows $\varepsilon$ to be used as a small parameter in a matrix perturbation analysis.

2.3. The Uni-directional Exit time $\tau_{A \rightarrow B}$. Following $[10, 14]$, we define a Markov random walk on our graph $(V, W)$ that jumps from node to node at unit time intervals. We define the probability to jump from node $i$ to node $j$ as $p_{i,j} = W_{i,j}/D_{i,i}$, where $D_{i,i} = \sum_j W_{i,j}$. For our purposes, it will also be useful to consider random walks on subsets of $V$, such as a random walk constrained to the cluster $A$, as well as random walks where some of the nodes are absorbing.

In particular, for the uni-directional exit time from cluster $A$ to cluster $B$, we consider a simplification of the weighted graph $(V, W)$, in which the cluster $B$ is collapsed into a new, single absorbing state $b$, with no allowed transitions $b \rightarrow A$. Let $V^{A,b} = A \uplus b$ be the node set of this simplified graph whose corresponding affinity matrix is

$$W^{A,b} = \begin{pmatrix}
W_{1,1} & \cdots & W_{1,k} & W_{1,b} \\
\vdots & \ddots & \vdots & \vdots \\
W_{k,1} & \cdots & W_{k,k} & W_{k,b} \\
0 & \cdots & 0 & 1
\end{pmatrix}.$$ 

Consider a random walk on this simplified graph, $(V^{A,b}, W^{A,b})$, whose corresponding Markov matrix is $(D^{A,b})^{-1} W^{A,b}$, where $D^{A,b}$ is the row-normalization diagonal matrix, $D_{i,i}^{A,b} = \sum_j W_{i,j}^{A,b}$. Denote the eigenvalues of this matrix by

$$1 = \lambda_1^{A,b} \geq \lambda_2^{A,b} \geq \ldots \geq \lambda_{k+1}^{A,b}.$$
Assume that the set $A$ is connected, so that $1 > \lambda_{2}^{A \cup B}$, and moreover that $\lambda_{2}^{A \cup B}$ is the second largest eigenvalue in absolute value. The characteristic exit time from $A$ to $B$ is defined as the relaxation time of this chain,

$$\tau_{A \to B} = \frac{1}{1 - \lambda_{2}^{A \cup B}}.$$  \hfill (2.6)

3. Main result. Our main result is a relation between the exit time $\tau_{A \to B}$ and the normalized cut measure Eq. (2.1). To this end, it will be useful to consider in addition a random walk constrained to stay in the cluster $A$. Under the assumption that $A$ is connected, this chain is reversible. We denote by $\lambda_{2}^{A}$ its second largest eigenvalue and by

$$\Gamma^{A} = 1 - \lambda_{2}^{A}$$  \hfill (3.1)

its spectral gap. Again we assume that the second largest eigenvalue $\lambda_{2}^{A}$ is also the second largest in absolute value.

We are finally ready to state our main result.

**Theorem 3.1.** Let $G$ be a symmetric graph with affinity matrix $W$. Consider a binary partition of the graph $V = A \sqcup B$. Assume that the subset $A$ is connected and that for the random walk constrained to remain in $A$, the second largest eigenvalue $\lambda_{2}^{A}$ is also second largest in absolute value. Denote by $\varepsilon = \max_{i \in A} W_{i,B} W_{i,A}$ the connectivity parameter.

1. For any $\varepsilon \geq 0$,

$$\frac{1}{\tau_{A \to B}} \leq N_{\text{cut}}(A, B).$$  \hfill (3.2)

2. Assume that

$$0 < \varepsilon < \frac{\Gamma^{A}}{8}.$$  \hfill (3.3)

Then

$$N_{\text{cut}}(A, B) = \frac{1}{\tau_{A \to B}} + \varepsilon^{2}\mathcal{R}(\varepsilon),$$  \hfill (3.4)

where

$$0 \leq \mathcal{R}(\varepsilon) \leq \frac{16}{\Gamma^{A}}.$$  

**Corollary 3.1.** Denote $\varepsilon = \max \left\{ \max_{i \in A} \frac{W_{i,B}}{W_{i,A}}, \max_{i \in B} \frac{W_{i,A}}{W_{i,B}} \right\}$ and assume that

$$0 < \varepsilon < \frac{1}{8} \min \left\{ \Gamma^{A}, \Gamma^{B} \right\}.$$  

Then

$$\text{MNcut}(A, B) = \frac{1}{\tau_{A \to B}} + \frac{1}{\tau_{B \to A}} + \varepsilon^{2}\mathcal{R}(\varepsilon),$$  \hfill (3.5)

where

$$0 \leq \mathcal{R}(\varepsilon) \leq \frac{16}{\Gamma^{A} + \Gamma^{B}}.$$
4. Discussion. For a binary partition, Eq. (3.5) gives a novel probabilistic interpretation to the normalized cut measure, in terms of characteristic exit times from each of the two subsets. For a partition into s disjoint subsets, \( V = \bigcup_{j=1}^{s} A_j \), Eq. (3.4) generalizes to a relation between the multiway normalized cut measure and the exit times from each of the s clusters,

\[
MNcut(\mathcal{A}_j) = \sum_{j=1}^{s} \frac{1}{\tau_{A_j \to A_j^c}} + \varepsilon^2 \mathcal{R}(\varepsilon)
\]

where \( A_j^c = V \setminus A_j \) and a remainder term \( \mathcal{R}(\varepsilon) \geq 0 \) with a bound analogous to the one in the binary case.

4.1. Limitations. The key assumption limiting the applicability of our result is Eq. (3.3), which requires that \( \varepsilon = \max_{i,j} W^{i,j} / W^{i,i} < (1-\lambda_2^A)/8 \). Hence, our result holds only when A and B are weakly connected, compared to the internal spectral gap of a random walk constrained to remain in A.

In addition, while a small value of \( \varepsilon \) implies a small normalized cut, since \( Ncut(A,B) \leq \varepsilon \), the converse is in general false: it is possible to construct examples where the normalized cut is small yet \( \varepsilon \) is large. Indeed \( \varepsilon \) is a “worse-case”, rather than an “on-average” measure of separation between clusters. Informally, a small value of \( \varepsilon \) means that the boundary between the clusters A and B is pronounced, in that each node in A has much stronger overall connection to nodes in A than to nodes in B.

4.2. Sharpness of the upper bound. A closer inspection of the proof shows that it in fact gives a stronger bound on the residual \( \mathcal{R}(\varepsilon) \) from Eq. (3.4).

To this end, consider a Markov chain \( X_t \) on \( A \cup b \) that starts from the stationary distribution \( \pi \) inside A,

\[
\pi_i = \begin{cases} 
W^{i,A}/W^{A,A} & i \in A \\
0 & i = b 
\end{cases}
\]

Then the one-step transition probability to the absorbing state b is

\[
P_{A \to B} = \sum_{i \in A} P(X_1 = b | X_0 = i) \pi_i = \sum_{i \in A} \left[ \frac{W^{i,B}}{W^{i,A} + W^{i,B}} \right] \frac{W^{i,A}}{W^{A,A}}.
\]

To factor out the small connectivity magnitude \( \varepsilon \), we write \( P_{A \to B} = \varepsilon Q_{A \to B} \). By definition, this quantity satisfies that \( Q_{A \to B} \leq 1 \) and measures the one-step absorption probability in units of \( \varepsilon \), when the Markov chain on A is in its stationary distribution. Then an improved bound on the residual is

\[
|R(\varepsilon)| \leq \frac{16}{\Gamma^A} \min \{2Q_{A \to B}, 1\}.
\]

Finally, let us remark that the coefficient 16 above is by no means sharp. For example, in the numerical experiment described below, the actual residual observed was approximately \( \frac{0.2}{\Gamma^A} \varepsilon^2 \).

4.3. Comparison to previous spectral bounds. Several authors derived the following lower bound on the multiway normalized cut of a partition into s clusters \( V = \bigcup_{j=1}^{s} A_j \):

\[
MNcut(\mathcal{A}_j) \geq \sum_{j=1}^{s} (1 - \lambda_j(V)),
\]

where \( \lambda_j(V) \) are the eigenvalues of the global random walk matrix on the whole graph \( G \) [8, 2].

In contrast, our result Eq. (4.1) is different as it provides (under suitable cluster separability conditions) both a lower and a precise upper bound on \( MNcut(V) \) in terms of local properties of the graph: the between cluster connectivity, internal relaxation times within each cluster, and the characteristic exit times from them. These quantities depend on eigenvalues of various normalized submatrices of G, which are in general different from the global eigenvalues \( \lambda_j(V) \).
4.4. Insensitivity to internal structure. One implication of our result is that when $A$ and $B$ are weakly connected, even though $Ncut(A, B)$ contains a normalization of the cut by the volume of a cluster, it nonetheless considers mainly the exit time from $A$, without taking into account its internal characteristics. For example, the set $A$ may in fact contain a disconnected or weakly connected subgraph, but this structure is not captured by the normalized cut measure. As discussed in [12], this leads to some limitations of the normalized cut measure in discovering clusters in graphs.

4.5. Approximation of exit times by Ncut. In the study of weakly coupled stochastic dynamical systems, a quantity of interest is the characteristic exit time from a given set of nodes. At low temperatures corresponding to extremely small values of $\varepsilon$, numerical methods for calculating $\lambda_2^{A\cup b}$ and hence $\tau_{A\to B}$ may be unstable as $\lambda_2^{A\cup b}$ becomes very close to 1. In contrast, the $Ncut$ measure can still be calculated efficiently. Our result thus provides a-priori guarantees for the accuracy of this approximation to $\lambda_2^{A\cup b}$. Similarly, the $Ncut$ approximation can be computed for extremely large graphs, for which eigenvalue computations may be quite challenging or even no longer feasible.

4.6. Analogy to continuum diffusion. Finally, we mention that our result can also be viewed as a discrete analog of known results regarding the second eigenvalue for a continuum diffusion process with two metastable states [9]. Consider a particle performing diffusion in a multi-dimensional potential with two wells separated by a high potential barrier. In this case, asymptotically as temperature tends to zero (equivalently in our case $\varepsilon \to 0$), the second (non-trivial) eigenvalue $\mu_2$ of the Fokker-Planck diffusion operator with two metastable states $A, B$ corresponding to the two potential wells is approximately given by

$$\mu_2 \approx \frac{1}{\tau_{A\to B}} + \frac{1}{\tau_{B\to A}}.$$

In analogy, this suggests that when there are two well defined clusters in a discrete finite graph, then the second eigenvalue of the whole graph is close to Eq. (1.1). A detailed study of this issue, which is related to isoperimetric and Cheeger inequalities on graphs, is an interesting problem beyond the scope of this paper.

5. A Numerical Example. One way to evaluate the tightness of the upper bound (3.4) is to compare the residual $Ncut(A, B) - 1/\tau_{A\to B}$ with our upper bound of $\frac{\lambda_2}{\varepsilon^2}$ in a numerical simulation. Unfortunately, as noted in Section 4.5, general-purpose numerical procedures do not allow us to measure $\lambda_2^{A\cup b}$ when $\varepsilon \ll 1$. We now construct a simple class of examples where $\lambda_2^{A\cup b}$ can be calculated analytically, allowing us to test the true residual against our upper bound, for values of $\varepsilon$ close to machine precision.

Assume that the cluster $A$ is a $d$-regular graph on $k$ nodes, drawn from the uniform distribution, so that the spectral gap $\Gamma^A$ is reasonably large. When all vertices of $A$ have the same exit probability $\alpha$ to the absorbing state $b$, it is easy to see that Theorem 3.1 holds with no approximation error, namely $Ncut(A, b) = 1 - \lambda_2^{A\cup b} = 1/\tau_{A\to b}$.

The simplest example where the exit probabilities are not all equal is when they are equal to $\alpha$ on one half of $A$ and to $\alpha r$ on the other half of $A$, for some $r > 0$. Consider then two $d$-regular graphs $(V^{(i)}, W^{(i)})$, $i = 1, 2$, sampled from the uniform distribution on $d$-regular graphs. Denote $V^{(1)} = \{1, \ldots, k\}$ and $V^{(2)} = \{k + 1, \ldots, 2k\}$, and let $b = 2k + 1$ be the absorbing state. Define an affinity matrix $W$ on $V = \{1, \ldots, 2k + 1\}$ by

$$W_{i,j} = \begin{cases} 1 & W_{i,j}^{(1)} = 1 \text{ or } W_{i,j}^{(2)} = 1 \text{ or } |i - j| = k \\ \alpha & 1 \leq i \leq k, j = 2k + 1 \\ r\alpha & k + 1 \leq i \leq 2k, j = 2k + 1 \end{cases}$$

so that each node in $V^{(1)}$ is connected to a corresponding node in $V^{(2)}$, to its neighbors in $V^{(1)}$, and to the absorbing state. Note that $\varepsilon = \alpha r/(d + 1)$ and that $\lambda_2^{A\cup b}(\alpha)$ is the second...
Fig. 5.1. Numerical Comparison of the normalized cut and the inverse exit time approximation

eigenvalue of the $3 \times 3$ Markov matrix

$$
\begin{pmatrix}
\frac{d}{d+1+\alpha} & \frac{1}{d+1+\alpha} & \frac{\alpha}{d+1+\alpha} \\
\frac{d+1+\alpha}{d+1+\rho} & \frac{d+1+\rho}{d+1+\rho} & \frac{d+1+\rho}{d+1+\rho} \\
0 & 0 & 1
\end{pmatrix},
$$

allowing us to calculate it analytically.

Figure 5 summarizes the result of this experiment for $k = 50$, $d = 7$, $r = 10$, and $\varepsilon \in [10^{-6}, 10^{-2}]$. We compared the exact value of the residual $Ncut(A, b) - \frac{1}{\tau_{A \to b}} = Ncut(A, b) - (1 - \lambda_2^{A,B})$ with the upper bound $\frac{16}{\Gamma_A} \varepsilon^2$. In the case under study, the upper bound is informative yet pessimistic (left panel), as the approximation $Ncut \approx 1 - \lambda_2^{A,B}$ is quite exact. The true residual $Ncut(A, b) - (1 - \lambda_2^{A,B})$ is in this case exactly quadratic in $\varepsilon$ (middle panel, log scale) and our quadratic bound is pessimistic by a multiplicative factor of roughly 80. Replacing the constant 16 in our main result, which is presumably not tight due to limitations of our proof technique, with the constant 1, we also compared the residual $Ncut(A, b) - (1 - \lambda_2^{A,B})$ with the quantity $\varepsilon^2 / \Gamma_A$ (right panel). The reader is invited to re-create this experiment, replacing the values $k, d, r$ and $\alpha$ with arbitrary values. The Matlab script and the original 7-regular graph used here are available at the author homepage http://gavish.web.stanford.edu

6. Proof of the main result. To simplify the proof, we first introduce the following notation: We write $d_{in}$ for the vector $(W_{i,A})_{i \in A}$ and $d_{out}$ for the vector $\varepsilon^{-1} (W_{i,B})_{i \in A}$. Hence $d_{in}(i)$ is the in-degree of a node $i \in A$, whereas $d_{out}(i)$ is its out-degree scaled by the connectivity parameter $\varepsilon$, so that $d_{out}(i) = O(1)$. Similarly, we define $d_{in} = \sum_{i \in A} d_{in}(i) = W_{A,A}$ and $d_{out} = \sum_{i \in A} d_{out}(i) = \varepsilon^{-1} W_{A,B}$. Our main object of interest, $\lambda_2^{A,B}$, will be denoted simply by $\lambda_2$, whereas the $Ncut$ measure is now given by

$$Ncut(A, B) = \frac{W_{A,B}}{W_{A,A} + W_{A,B}} = \frac{\varepsilon d_{out}}{d_{in} + \varepsilon d_{out}}.$$
For future use, we denote $D_{in} = \text{diag}(d_{in}(1), \ldots, d_{in}(k))$, $D_{out} = \text{diag}(d_{out}(1), \ldots, d_{out}(k))$ and $D(\varepsilon) = D_{in} + \varepsilon D_{out}$, so that

$$D^{A:ib} = \begin{pmatrix} D_{in} + \varepsilon D_{out} & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} D(\varepsilon) & 0 \\ 0 & 1 \end{pmatrix}.$$  

Finally, let $W^A$ be the upper $k \times k$ minor of $W$, namely $W^A_{i,j} = W_{i,j}$, for $1 \leq i, j \leq k$.

6.1. Passing to a generalized eigenvalue problem. Recall that $\lambda^A_{2:ib}$ is the second largest root of the characteristic polynomial $\det((D^{A:ib})^{-1}W^{A:ib} - \lambda I) = 0$, or equivalently of $\det(W^{A:ib} - \lambda D^{A:ib}) = 0$. Expanding the latter determinant by the last row of $W^{A:ib}$ gives

$$\det(W^{A:ib} - \lambda D^{A:ib}) = \pm (1 - \lambda) \cdot \det(W^A - \lambda D(\varepsilon)).$$

Since $\lambda^A_{2:ib} < 1$, it is enough to consider the smaller $k \times k$ matrices $D(\varepsilon)$ and $W^A$. Obviously, $\mathbf{x} \neq 0$ is an eigenvector of $D(\varepsilon)^{-1}W^A$ corresponding to an eigenvalue $\lambda$ if and only if

$$W^A \mathbf{x} = \lambda D(\varepsilon) \mathbf{x}.$$  

(6.1)

When (6.1) holds, we say that $\lambda$ is a generalized eigenvalue of the matrix pair $(W^A, D(\varepsilon))$. Clearly, the solutions to (6.1) are $\lambda_2 \equiv \lambda^A_{2:ib} \geq \ldots \geq \lambda^A_{k+1}$.  

6.2. Lower bound for $\text{Ncut}(A, B)$. The lower bound (3.2) follows from a min-max characterization of generalized eigenvalues [18, pp. 281 part VI corollary 1.16], whereby the largest eigenvalue of the matrix pair $(W^A, D(\varepsilon))$ is given by

$$\lambda^A_{2:ib} = \max_{\mathbf{x} \neq 0} \frac{\mathbf{x}^\top W^A \mathbf{x}}{\mathbf{x}^\top D(\varepsilon) \mathbf{x}} \geq \frac{1^\top W^A 1}{1^\top D(\varepsilon) 1} = \frac{d_{in}}{d_{in} + \varepsilon d_{out}} = 1 - \text{Ncut}(A, B).$$  

(6.2)

We note that this simple inequality can also be derived by other methods. For example, from a lemma by Ky Fan [8, 2] it follows that $\text{Ncut}(A, B) \geq 2 - \lambda^A_{2:ib} - \lambda^A_{2:ib} = 1 - \lambda^A_{2:ib}$.

6.3. Upper bound. The proof of the upper bound relies on perturbation theory of generalized eigenvalues [18, ch. 6], where we treat matrix pair above as a perturbed pair:

$$(W^A, D(\varepsilon)) = (W^A, D_{in}) + \varepsilon (0, D_{out}).$$

Two facts simplify the perturbation analysis to follow. First, the perturbation does not affect the left matrix $W^A$. Second, the right hand side matrices on both the perturbed and unperturbed pair are diagonal.

We will need the following key lemma, which is proved in the Appendix. Recall the notation $Q_{A\rightarrow B} = \varepsilon^{-1}P_{A\rightarrow B}$, as defined in (4.2).

**Lemma 6.1.** If $\varepsilon < \frac{1 - \text{Ncut}(A, B)}{8}$, then there exists $\beta \in \mathbb{R}$ with $|\beta| \leq \frac{8}{1 - \text{Ncut}(A, B)} \min\{2\text{Ncut}(A, B), 1\}$ such that

$$\Lambda = \frac{1}{1 + \varepsilon \frac{d_{out}}{d_{in}} + \varepsilon^2 \beta}$$  

(6.3)

is a generalized eigenvalue of the matrix pair $(W^A, D(\varepsilon))$.

Given our assumption that $\varepsilon$ satisfies (3.3), Lemma 6.1 holds. Hence let $\beta$ be such that $\Lambda = \left(1 + \varepsilon \frac{d_{out}}{d_{in}} + \varepsilon^2 \beta\right)^{-1}$ is a generalized eigenvalue of the matrix pair $(W^A, D(\varepsilon))$. We now conclude the proof in two steps:

1. We show that $\text{Ncut}(A, B) = 1 - \Lambda + \varepsilon^2 \mathcal{R}(\varepsilon)$.
2. We then show that $\Lambda$ must in fact be this pair’s largest eigenvalue, so that necessarily $\Lambda = \lambda_2$.

From these two steps the theorem readily follows.

For the first step, we rewrite (6.3) as

$$\Lambda = \frac{d_{in}}{1 + \varepsilon d_{out}} \left( \frac{1}{1 + \varepsilon^2 \beta \cdot \frac{d_{in}}{d_{in} + \varepsilon d_{out}}} \right).$$

(6.4)

Now, for any $\delta \in \left[ -\frac{1}{2}, \frac{1}{2} \right]$ we have $1 + \delta \leq 1 + 2|\delta|$. As $\varepsilon < \frac{\Gamma_A}{8}$ and $|\beta| \leq \frac{8}{16}$, obviously

$$\varepsilon^2 |\beta| \cdot \frac{d_{in}}{d_{in} + \varepsilon d_{out}} \leq \left( \frac{\Gamma_A}{8} \right)^2 |\beta| \leq \frac{1}{2}.$$  

Hence,

$$\frac{1}{1 + \varepsilon^2 \beta \cdot \frac{d_{in}}{d_{in} + \varepsilon d_{out}}} \leq 1 + 2\varepsilon^2 |\beta| \cdot \frac{d_{in}}{d_{in} + \varepsilon d_{out}} \leq 1 + 2\varepsilon^2 |\beta|.$$  

(6.5)

Combining (6.4) and (6.5) gives that

$$\Lambda = \frac{1}{1 + \varepsilon \cdot \frac{d_{out}}{d_{in}} + \varepsilon^2 \beta} = \frac{d_{in}}{d_{in} + \varepsilon d_{out}} + \varepsilon^2 \mathcal{R}(\varepsilon) = 1 - \text{Ncut}(A,B) + \varepsilon^2 \mathcal{R}(\varepsilon)$$

with

$$|\mathcal{R}(\varepsilon)| \leq \frac{16}{\Gamma_A} \min \{ 2\mathcal{Q}_{A\rightarrow B}, 1 \}$$

as required.

For the second step, we need to show that if $\varepsilon < \frac{\Gamma_A}{8}$, then the eigenvalue $\Lambda$ is indeed the largest eigenvalue, $\lambda_2$. This is achieved by proving that $\lambda_{A,lb}$, the next largest eigenvalue of $(W^A, D(\varepsilon))$, is strictly smaller than $\Lambda$. To this end, we appeal to [18, pp. 314 part VI theorem 3.2] whereby if

$$\zeta = \varepsilon \cdot \max_{\|x\|_2=1} \frac{x^TD_{out}x}{\sqrt{(x^TW^Ax)^2 + (x^TD_{in}x)^2}} < 1,$$

(6.6)

then

$$\frac{|\lambda_{A,lb} - \lambda_2|}{\sqrt{(\lambda_2)^2 + 1} \cdot \sqrt{(\lambda_{A,lb})^2 + 1}} \leq \zeta.$$  

(6.7)

Here as before, $\lambda_2^A$ is the second eigenvalue of $(W^A, D(0))$. To see that the condition $\zeta < 1$ holds, observe that from (6.6) we have

$$\zeta \leq \varepsilon \cdot \max_{\|x\|_2=1} \frac{x^TD_{out}x}{x^TD_{in}x} = \varepsilon \cdot \max_i \frac{d_{out}(i)}{d_{in}(i)} = \varepsilon < 1.$$

We thus deduce from (6.7) that

$$\frac{|\lambda_{A,lb} - \lambda_2^A|}{2} \leq \frac{|\lambda_{A,lb} - \lambda_2^A|}{\sqrt{(\lambda_2^A)^2 + 1} \cdot \sqrt{(\lambda_{A,lb})^2 + 1}} \leq \zeta \leq \varepsilon.$$
and in particular
\[
\lambda_3^{A,ub} \leq \lambda_2^A + 2\varepsilon. \tag{6.8}
\]

To complete the second step, as illustrated in Figure 6.1, we now localize \( \Lambda \) and \( \lambda_3^{A,ub} \) in two disjoint intervals, which implies that necessarily \( \lambda_3^{A,ub} < \Lambda \). To this end, first observe that for all \( \delta \in [-\frac{1}{2}, \frac{1}{2}] \), we have \( 1 - |\delta| \leq \frac{1}{1+\varepsilon} \). Similarly to (6.5), this implies that
\[
\Lambda = d_{in} \left( \frac{1}{1+\varepsilon^2 |\beta|} \cdot \frac{d_{in}}{d_{in} + \varepsilon d_{out}} \right) \geq \frac{d_{in}}{d_{in} + \varepsilon d_{out}} \left( 1 - \varepsilon^2 |\beta| \cdot \frac{d_{in}}{d_{in} + \varepsilon d_{out}} \right),
\]
so that \( \Lambda \geq \frac{d_{in}}{d_{in} + \varepsilon d_{out}} - \varepsilon^2 |\beta| \). Now, clearly
\[
\frac{d_{in}}{d_{in} + \varepsilon d_{out}} = \frac{1}{1 + \varepsilon \frac{d_{out}}{d_{in}}} \geq 1 - \varepsilon \frac{d_{out}}{d_{in}} \geq 1 - \varepsilon.
\]
Finally, since \( |\beta| \leq \frac{8}{\Gamma} \) and \( \varepsilon < \frac{\Gamma}{8} \),
\[
\Lambda \geq \frac{d_{in}}{d_{in} + \varepsilon d_{out}} - \varepsilon^2 |\beta| \geq 1 - \varepsilon - \varepsilon = 1 - 2\varepsilon. \tag{6.9}
\]
Combining (6.9), (6.8) and the fact that \( \Gamma^A = 1 - \lambda_2^A \geq 8\varepsilon \), we obtain
\[
\lambda_3^{A,ub} \leq 1 - \Gamma^A + 2\varepsilon \leq 1 - 8\varepsilon + 2\varepsilon = 1 - 6\varepsilon < 1 - 2\varepsilon \leq \Lambda.
\]
Hence, the eigenvalue \( \Lambda \) is strictly larger than \( \lambda_3^{A,ub} \), namely \( \Lambda = \lambda_2^{A,ub} \). The requested upper bound and Theorem 3.1 follow. \( \square \)

Acknowledgements. We thank the anonymous referees for their helpful comments and suggestions. MG is supported by a William R. and Sara Hart Kimball Stanford Graduate Fellowship. BN’s research was partly supported by a grant from the ISF.

Appendix A.

Lemma 6.1 finds an eigenvalue for the perturbed matrix \((W^A, D(\varepsilon)) = (W^A, D_{in}) + \varepsilon (0, D_{out})\). To prove the lemma, following [18], we introduce a tool to study the effect of a diagonal perturbation on generalized eigenvalues.
Let the spectrum of the reversible Markov matrix $D_{in}^{-1}W_A$ be
\[ 1 = \lambda_1^A > \lambda_2^A \geq \lambda_3^A \geq \cdots \geq \lambda_k^A, \]
and define
\[ A_2 = \begin{pmatrix} \lambda_2^A & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_k^A \end{pmatrix}. \]

 Equip the product space $\mathbb{R}^{k-1} \times \mathbb{R}^{k-1}$ with the norm
\[ \|(v, w)\|_F = \max \{\|v\|_2, \|w\|_2\} \]
where $v, w \in \mathbb{R}^{k-1}$ and $\|\cdot\|_2$ is the usual Euclidean norm.

We now define an operator $T : \mathbb{R}^{k-1} \times \mathbb{R}^{k-1} \to \mathbb{R}^{k-1} \times \mathbb{R}^{k-1}$ by
\[ T(v, w) = (w + A_2v, w + v), \]
and its associated quantity $\text{dif}[A_2]$ by
\[ \text{dif}[A_2] = \inf_{\|(v, w)\|_F = 1} \|T(v, w)\|_F. \]

The following auxiliary lemma relates this quantity to the spectral gap of $A$.

**Lemma A.1.** Assume that $\lambda_2^A = \max_{j \geq 2} |\lambda_j^A|$. Then,
\[ \text{dif}[A_2] = \frac{1 - \lambda_2^A}{2} = \frac{\Gamma^A}{2}. \]

**Proof.** Let $(v, w)$ be a pair with $\|(v, w)\|_F = 1$. Therefore, either $\|w\|_2 = 1$ or $\|v\|_2 = 1$.
When $\|w\|_2 = 1$,
\[ \|T(v, w)\|_F \geq \|w + A_2v\|_2 \geq \|w\|_2 - \|A_2v\|_2. \]
Since $\|v\|_2 \leq 1$ and $\|A_2\| = \lambda_2^A$, it follows that
\[ \|T(v, w)\|_F \geq 1 - \lambda_2^A = \Gamma^A. \]
Next, consider the case where $\|v\|_2 = 1$. Then by definition $\|A_2v\|_2 \leq \lambda_2^A < 1$. It is easy to see that inside the unit ball $\|w\|_2 \leq 1$, the minimizer of $\|T(v, w)\|$ is simply the average $w = -\frac{1}{2}(v + A_2v)$, which gives
\[ \|T(v, w)\|_F = \|v - A_2v\|_2/2. \]
To conclude the proof we note that since $A_2$ is a diagonal matrix with largest entry $\lambda_2^A$ in position $(1, 1)$, the quantity $\|v - A_2v\|$ is minimized at the vector $v = e_1$. In this case
\[ \frac{\|e_1 - A_2e_1\|_2}{2} = \frac{1 - \lambda_2^A}{2} = \frac{\Gamma^A}{2}. \]

**Proof of Lemma 6.1:** Recall that the matrix pair $(W^A, D(\varepsilon))$ is a sum of the pair $(W^A, D_{in})$ and a diagonal perturbation:
\[ (W^A, D(\varepsilon)) = (W^A, D_{in}) + \varepsilon (0, D_{out}). \]
For the unperturbed pair, observe that its generalized eigenvalues are all real, since the matrix \( D_{in}^{-1} W \) is similar to the symmetric matrix \( D_{in}^{-1/2} (D_{in}^{-1} W A) D_{in}^{-1/2} = D_{in}^{-1/2} W A D_{in}^{-1/2} \). Furthermore, since \( D_{in}^{-1} W \) is the transition matrix of an irreducible Markov chain, the largest eigenvalue in (A.1) is \( \lambda_1^A = 1 \) and the remaining eigenvalues are all strictly smaller than 1 in absolute value.

Now, the vector of length \( k = \frac{1}{\sqrt{d_{in}}} (1, \ldots, 1)^T \) is an eigenvector of \( D_{in}^{-1/2} W A D_{in}^{-1/2} \) corresponding to the eigenvalue \( \lambda_1^A = 1 \). Let

\[
U = \begin{pmatrix}
\mathbf{x} & \mathbf{u}_1 & \cdots & \mathbf{u}_{k-1}
\end{pmatrix}
\]

be the \( k \times k \) orthogonal matrix that diagonalizes \( D_{in}^{-1/2} W A D_{in}^{-1/2} \), namely

\[
D_{in}^{-1/2} W A D_{in}^{-1/2} = U \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & \lambda_2^A & 0 & 0 \\
0 & 0 & \ddots & 0 \\
0 & 0 & 0 & \lambda_k^A
\end{pmatrix} U^T.
\]

Now define \( X = D_{in}^{-1/2} U \) and decompose \( X \) to blocks as \( X = (X_1, X_2) \) where

\[
X_1 = \frac{1}{\sqrt{d_{in}}} (1, \ldots, 1)^T
\]

is a \( k \times 1 \) vector and \( X_2 \) is the remaining \( k \times (k-1) \) matrix,

\[
X_2 = D_{in}^{-1/2} \begin{pmatrix}
\mathbf{u}_1 & \cdots & \mathbf{u}_{k-1}
\end{pmatrix}
\]

Clearly we have

\[
X^T W A X = \begin{pmatrix}
1 & \lambda_2^A \\
& \ddots & \ddots \\
& & \lambda_k^A
\end{pmatrix}
\]

and \( X^T D_{in} X = I_k \). In matrix block notation,

\[
\begin{pmatrix}
X_1^T \\
X_2^T
\end{pmatrix} \cdot W A \cdot (X_1, X_2) = \begin{pmatrix}
1 & 0 \\
0 & A_2
\end{pmatrix}
\]

and

\[
\begin{pmatrix}
X_1^T \\
X_2^T
\end{pmatrix} \cdot D_{in} \cdot (X_1, X_2) = \begin{pmatrix}
1 & 0 \\
0 & I_{k-1}
\end{pmatrix}.
\]

With this simultaneous diagonalization of the unperturbed matrix pair \((W A, D_{in})\) at hand, we turn to the perturbation pair, \((0, D_{out})\). Again using matrix block notation, denote

\[
\begin{pmatrix}
X_1^T \\
X_2^T
\end{pmatrix} \cdot D_{out} \cdot (X_1, X_2) = \begin{pmatrix}
\mathbf{d}_{out}^{12} & \mathbf{d}_{out}^{12} \\
\mathbf{d}_{out}^{12} & \mathbf{d}_{out}^{12}
\end{pmatrix},
\]

12
where \( \mathbf{d}_{out}^{12} \) is a \( 1 \times (k-1) \) vector, \( \mathbf{d}_{out}^{21} = (\mathbf{d}_{out}^{12})^\top \) and \( D_{out}^{22} \) is a \( (k-1) \times (k-1) \) matrix. It is straightforward to verify that

\[
\mathbf{d}_{out}^{12} = \frac{1}{\sqrt{d_{in}}} \left( \frac{\mathbf{d}_{out}(1)}{\sqrt{d_{in}(1)}}, \cdots, \frac{\mathbf{d}_{out}(k)}{\sqrt{d_{in}(k)}} \right) \left( \begin{array}{c} \mathbf{u}_1 \vline \cdots \vline \mathbf{u}_{k-1} \end{array} \right). \tag{A.4}
\]

The amount by which the eigenvalue \( \lambda_2^A \) is perturbed as \( \varepsilon \) departs from 0 depends primarily on the norms of \( \mathbf{d}_{out}^{12} \) and \( D_{out}^{22} \). For \( \mathbf{d}_{out}^{12} \) we have that

\[
\|\mathbf{d}_{out}^{12}\|_2^2 \leq \frac{1}{d_{in}} \sum_{i=1}^k \frac{d_{out}^2(i)}{d_{in}(i)} = \sum_{i=1}^k \frac{d_{out}^2(i)}{d_{in}(i)} \cdot \frac{d_{in}(i)}{d_{in}} = \sum_{i=1}^k \frac{d_{out}^2(i)}{d_{in}^2(i)} \pi^A(i),
\]

where \( \pi^A(i) = \frac{d_{in}(i)}{\sum_{i=1}^k d_{in}(i)} \) is the stationary distribution of a random walk inside \( A \). Furthermore, since \( \max_i \frac{d_{out}(i)}{d_{in}(i)} = 1 \), it follows that for all \( i \in A, d_{in}(i) \geq \frac{1}{2} [d_{in}(i) + \varepsilon d_{out}(i)] \). Hence,

\[
\|\mathbf{d}_{out}^{12}\|_2^2 \leq \sum_{i=1}^k \frac{d_{out}^2(i)}{d_{in}(i)} \cdot \pi^A(i) \leq \sum_{i=1}^k \frac{2d_{out}(i)}{d_{in}(i)} + \varepsilon d_{out}(i) \cdot \pi^A(i) = 2Q_{A \rightarrow B}.
\]

Since by (A.4) clearly \( \|\mathbf{d}_{out}^{12}\|_2^2 \leq 1 \), we can write

\[
\|\mathbf{d}_{out}^{12}\|_2^2 \leq \min \{2Q_{A \rightarrow B}, 1\}. \tag{A.5}
\]

As for \( D_{out}^{22} \), it is easy to verify that since \( U \) is orthogonal, \( \|D_{out}^{22}\|_F \leq 1 \).

Now, as the generalized eigenvalues of the pair \((A_2, \mathbf{I}_{k-1})\) are strictly smaller than 1 in absolute value, the column \( X_1 \) spans a simple eigenspace of the unperturbed pair \((W^A, D_{in})\). We can thus appeal to the main instrument in our proof, a theorem from matrix perturbation theory for generalized eigenvalue problems [18, pp. 310 part VI theorem 2.15].

In our setting and our notation, this theorem states that if \( \text{dif} [A_2] > 0 \) and if

\[
\frac{\varepsilon \gamma}{\text{dif} [A_2] - \varepsilon \rho} < \frac{1}{2}, \tag{A.6}
\]

where

\[
\gamma = \|\mathbf{d}_{out}^{12}\|_2 = \|\mathbf{d}_{out}^{21}\|_2, \quad \rho = \frac{d_{out}}{d_{in}} + \|D_{out}^{22}\|_F,
\]

then there exists \( \mathbf{p} \in \mathbb{R}^{k-1} \) with

\[
\|\mathbf{p}\|_2 \leq 2 \cdot \frac{\gamma}{\text{dif} [A_2] - \varepsilon \rho}, \tag{A.7}
\]

such that the pair

\[
\left( 1, \left( 1 + \frac{\varepsilon d_{out}}{d_{in}} + \varepsilon^2 d_{out}^{12} \cdot \mathbf{p} \right)^{-1} \right)
\]

is a generalized eigenvalue of the matrix pair \((W^A, D(\varepsilon))\).

The first condition in this theorem, namely \( \text{dif} [A_2] > 0 \), follows directly from Lemma A.1. For the second condition, we need to bound the constants \( \gamma \) and \( \rho \), as follows. For \( \gamma \), we have seen in (A.5) that \( \gamma = \|\mathbf{d}_{out}^{21}\|_2 = \|\mathbf{d}_{out}^{12}\|_2 \leq 1 \). Similarly, we have \( \rho = \frac{d_{out}}{d_{in}} + \|D_{out}^{22}\|_F \leq 2. \)
Suppose now that $\varepsilon < \frac{\Gamma_A^4}{4}$. Then we have $\varepsilon \rho < \frac{\Gamma_A^4}{4}$. Invoking Lemma A.1, whereby $\text{dif} [A_2] = \frac{\Gamma_A^4}{2}$, we get that

$$\text{dif} [A_2] - \varepsilon \rho > \frac{\Gamma_A^4}{2} - \frac{\Gamma_A^4}{4} = \frac{\Gamma_A^4}{4}$$ \hspace{1cm} \text{(A.8)}$$

and therefore

$$\frac{\varepsilon \gamma}{\text{dif} [A_2] - \varepsilon \rho} < \frac{\varepsilon \gamma}{\frac{\Gamma_A^4}{4}} < \frac{\Gamma_A^4}{\frac{\Gamma_A^4}{4}} = \frac{1}{2}$$

so the second condition of the theorem, condition (A.6), is satisfied.

In conclusion, the theorem above holds, so a vector $p$ as above exists, with

$$\|p\|_2 \leq \frac{2 \gamma}{\text{dif} [A_2] - \varepsilon \rho}.$$ \hspace{1cm} \text{(A.9)}$$

Finally, using (A.9), (A.8) and (A.5) and the definition $\gamma = \|d_{12}^{12}\|_2$, we have

$$|d_{12}^{12} \cdot p| \leq \|p\|_2 \|d_{12}^{12}\|_2 \leq \frac{2 \gamma}{\text{dif} [A_2] - \varepsilon \rho} \|d_{12}^{12}\|_2 = \frac{2}{\text{dif} [A_2] - \varepsilon \rho} \frac{\|d_{12}^{12}\|_2^2}{\text{dif} [A_2] - \varepsilon \rho} \leq \frac{2 \|d_{12}^{12}\|_2^2}{\frac{\Gamma_A^4}{4}} \leq \frac{8}{\Gamma_A^4} \min \{2Q_{A \rightarrow B}, 1\}.$$ 

This completes the proof of Lemma 6.1. \hfill \Box

REFERENCES


